UNITED STATES PATENT APPLICATION FOR

DIGITAL RECORDING AND PLAYBACK SYSTEM
WITH VOICE RECOGNITION CAPABILITY
FOR CONCURRENT TEXT GENERATION

Inventor:

Takashi Nakatsuyama

Prepared by:

WAGNER, MURABITO & HAO LLP

TWO NORTH MARKET STREET

THIRD FLOOR

SAN JOSE, CALIFORNIA 95113

(408) 938-9060

Confidential

DIGITAL RECORDING AND PLAYBACK SYSTEM
WITH VOICE RECOGNITION CAPABILITY
FOR CONCURRENT TEXT GENERATION

RELATED U.S. APPLICATIONS

This application is a continuation of and claims priority to U.S. Patent No. 6,754,619 entitled "Digital Recording And Playback System With Voice Recognition Capability For Concurrent Text Generation," by Nakatsuyama filed on 11/15/99, which is incorporated herein by reference.

BACKGROUND OF THE INVENTION

FIELD OF THE INVENTION

The present invention relates to the design of digital recording and playback systems. More specifically, the present invention pertains to the processing of voice and concurrent generation of corresponding text in a portable digital appliance.

RELATED ART

The use of portable digital recording and playback devices are quickly gaining popularity in business and among individual users. In particular, one attractive feature of digital recording is the possibility of converting the voice messages into text, which can then be reviewed, revised and incorporated into documents or otherwise retrieved for use subsequently. Today, there are several models of portable digital recorder in the marketplace. These prior art recorders typically record voice messages as compressed digital data. In order to convert the compressed digital data to text data, a separate computer program

is generally required. Thus, in the prior art, subsequent to a recording session, the user needs to post-process the compressed digital data to perform the voice-to-text conversion. This requires additional processing time, and in some cases even requires the user to transfer the compressed digital data from the portable

5   device to a personal computer (PC) having the necessary software program before the conversion can be performed. It is desirable to eliminate the extra step of post-recording conversion from compressed digital data to text data in a portable digital recording and playback system.

10   These prior art devices are not well-suited for generating text data from the recorded voice data for an additional reason. In order to achieve good conversion from voice to text, a high quality voice input to the voice to text conversion engine is needed. In prior art portable systems, the voice data is subject to high compression because portable systems typically have limited

15   memory capacity, and high compression allows more voice data to be stored into the limited memory resources. Since voice data is stored in a highly compressed format in these portable prior art devices, the text data generated directly from the compressed voice data by a conversion program is usually unsatisfactory. As such, it is highly advantageous to have a portable digital recording and

20   playback system which provides high quality conversion from voice to text.

Furthermore, portable devices are typically battery-powered. Thus, the need to conserve power is a major design consideration. As such, while a high capacity stager can potentially be used in a large, non-portable device deriving

25   its power from a power outlet to improve the quality of the conversion from compressed voice data to text data, it is not a viable option in a portable device. Therefore, there exists a need for a portable digital recording and playback

system which provides high quality conversion from voice to text and yet does

not require a high rate of power consumption.

SUMMARY OF THE INVENTION

In implementing a viable portable digital recording and playback system, it is highly desirable that components that are well known in the art and are compatible with existing computer systems and other appliances be used so that

5   the cost of realizing the portable digital recording and playback system is low. By so doing, the need to incur costly expenditures for retrofitting existing computer systems and other appliances or for building custom components is advantageously eliminated.

10   Thus, a need exists for a portable digital recording and playback system which does not require post-recording conversion to generate text data from compressed digital data. A further need exists for a portable digital recording and playback system which meets the above need and which provides high quality conversion from voice to text. Still another need exists for a portable

15   digital recording and playback system which meets both of the above needs and which does not require a high level of power consumption. Yet another need exists for a portable digital recording and playback system which meets all of the above needs and which is conducive to use with existing computer systems and other appliances.

20

Accordingly, the present invention provides a portable digital recording and playback system which generates text data from voice without requiring post-recording conversion from compressed digital data to text data. The present invention further provides a portable digital recording and playback

25   system which not only provides voice to text conversion without post-processing but the conversion is also of high quality. Embodiments of the present invention perform voice-to-text conversion using the high quality audio input signal rather

than highly compressed voice data so that high quality conversion is achieved. Moreover, the present invention provides a portable digital recording and playback system which includes the above features and which conserves power for full battery operation. Furthermore, embodiments of the present invention

5 utilize components that are well known in the art and are compatible with existing computer systems and other appliances, so that the present invention is conducive for use with existing computer systems and other appliances. These and other advantages of the present invention not specifically mentioned above will become clear within discussions of the present invention presented herein.

10

More specifically, in one embodiment of the present invention, a digital recording and playback system is provided. In this embodiment, the system comprises an audio capturing device configured to receive a voice input. The system also comprises a high compression encoder (HCE) coupled to the audio

15 capturing device and configured to generate digital wave data corresponding to the voice input. The system further comprises a voice recognition engine (VRE) coupled to the audio capturing device and configured to generate text data corresponding to the voice input. Moreover, in this embodiment, the HCE and VRE are selectively coupled to a memory sub-system which is configured to

20 store the digital wave data and the text data. In particular, in this embodiment, the HCE and the VRE are operable to concurrently generate the digital wave data and the text data in response to the voice input such that the digital wave data and the text data can be stored in the memory sub-system in a synchronized manner. Thus, in this embodiment, the present invention provides

25 recording capability wherein text data is generated from a voice input without requiring post-recording conversion. In a specific embodiment, the present invention includes the above and wherein the system is battery-powered.

Additional embodiments of the present invention include the above and further comprise a decoder selectively coupled to the memory sub-system and configured to decode the digital wave data into decoded audio data, a digital-to-analog (D/A) converter coupled to the decoder and configured to convert the decoded audio data into an analog signal, and an audio output device coupled to the D/A converter and configured to generate a voice output corresponding to the voice input from the analog signal. Moreover, these embodiments also comprises a display sub-system selectively coupled to the memory sub-system and configured to display the text data. Thus, in these embodiments, the present invention provides simultaneous voice playback and text display.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated in and form a part of this specification, illustrate embodiments of the invention and, together with the

5   description, serve to explain the principles of the invention:

Figure 1 is a block diagram illustrating a portable digital recording and playback system 100 in accordance with one embodiment of the present invention, wherein the system has built-in voice recognition capability for

10   concurrent text generation during voice recording.

Figure 2A is a flow diagram illustrating steps for performing recording using system 100 of Figure 1 in accordance with one embodiment of the present invention.

15

Figure 2B is a diagram illustrating one embodiment of arrangement of corresponding portions of voice data and text data as stored in a portable digital recording and playback system 100 in accordance with the present invention.

20   Figure 3 is a flow diagram illustrating steps for performing playback using system 100 of Figure 1 in accordance with one embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

In the following detailed description of the present invention, a digital recording and playback system with voice recognition capability for concurrent text generation, numerous specific details are set forth in order to provide a

5   thorough understanding of the present invention. However, it will be recognized by one skilled in the art that the present invention may be practiced without these specific details or with equivalents thereof. In other instances, well known methods, procedures, components, and circuits have not been described in detail as not to unnecessarily obscure aspects of the present invention.

10

Exemplary Configuration of a Digital Recording and Playback System

of the Present Invention

Figure 1 is a block diagram illustrating a portable digital recording and playback system 100 in accordance with one embodiment of the present

15   invention, wherein the system has built-in voice recognition capability for concurrent text generation during voice recording. In system 100, an audio capturing device 110 is coupled to a high compression encoder (HCE) 120. Audio capturing device 110 is also coupled to a voice recognition engine (VRE) 130. Both HCE 120 and VRE 130 are selectively coupled to a memory sub-

20   system 140 through an intelligent switch 135. More particularly, switch 135 is operable to couple either HCE 120 or VRE 130, but not both, to memory sub-system 140 at any given time. In one embodiment, switch 135 is a multiplexer. In another embodiment, switch 135 is a software switch for data routing. In an exemplary embodiment, audio capturing device 110 comprises a microphone. It

Confidential

is appreciated that audio signals are supplied to HCE 120 and VRE 130 simultaneously so that voice encoding and recognition functions can be performed in parallel.

5   It is appreciated that within the scope of the present invention, memory sub-system 140 can comprise volatile memory (e.g., random access memory RAM), non-volatile memory (e.g., read only memory ROM), and/or data storage devices such as magnetic or optical disk drives and disks (e.g., diskettes, tapes, cartridges) which are computer readable media for storing information and

10   instructions. These memory modules of memory sub-system 140 can be removable to facilitate the easy transfer of data stored therein. In one embodiment, memory sub-system 140 comprises semiconductor flash memory.

Still referring to Figure 1, memory sub-system 140 is selectively coupled

15   to both a decoder 150 and a display sub-system 180 through an intelligent switch 145. More particularly, switch 145 is operable to couple memory sub-system 140 to either decoder 150 or display sub-system 180, but not both, at any given time. In one embodiment, switch 145 is a multiplexer. In another embodiment, switch 145 is a software switch. In one embodiment, switch 145 is controlled by

20   the texted voice data generated by VRE 130. Moreover, in an exemplary embodiment, display sub-system 180 comprises flat panel display technology, for example, a liquid crystal display (LCD).

With reference still to Figure 1, decoder 150 is further coupled to a digital-to-analog (D/A) converter 160. Moreover, D/A converter 160 is coupled to an amplifier 165, which is in turn coupled to an audio output device 170. In one embodiment, audio output device 170 comprises a speaker.

5

With reference still to Figure 1, in one embodiment, an editing sub-system 190 is coupled to memory sub-system 140. In this embodiment, editing sub-system 190 can include an alphanumeric input device having alphanumeric and function keys to allow user editing of the text data. Editing sub-system 190 can

10   also include a cursor control or directing device to facilitate text editing and command selection by a user. Cursor control device allows the computer user to dynamically signal the two dimensional movement of a visible symbol (cursor) on a screen of display sub-system 180. Many implementations of cursor control device are known in the art including a trackball, mouse, touch pad, joystick or

15   special keys on the alphanumeric input device capable of signaling movement of a given direction or manner of displacement. Alternatively, it will be appreciated that a cursor can be directed and/or activated via input from the alphanumeric input device using special keys and key sequence commands. The present invention is also well suited to directing a cursor by other means such as, for

20   example, voice commands. Moreover, editing sub-system 190 can further include a printing device for generating paper copies of the text data.

Operation of a Digital Recording and Playback System of the Present Invention

Referring next to Figure 2A, a flow diagram 200 illustrating steps for performing recording using system 100 of Figure 1 in accordance with one embodiment of the present invention is shown. In step 210, system 100 receives a voice input using audio capturing device 110.

5

In step 220, system 100 of Figure 1 generates digital wave data from the voice input using HCE 120. In an exemplary embodiment, HCE 120 of system 100 can achieve a compression ~~ratio~~ rate of two kilobits per second (2 kbit/s). It is appreciated that the high level of compression of the digital wave data in

10   accordance with the present invention advantageously reduces the amount of memory that is required to store the digital wave data.

In step 230, system 100 of Figure 1 generates text data from the voice input using VRE 130. In one embodiment, VRE 130 of system 100 uses Hidden

15   Markov Model (HMM) techniques to perform voice recognition, although other voice recognition techniques can also be used within the scope of the present invention. It is also appreciated that the text data can be in any of a wide variety of formats. In an exemplary embodiment, the text data is generated in hypertext markup language (HTML) format.

20

Referring still to Figure 2A, in step 240, system 100 of Figure 1 stores the digital wave data and the text data as mixed data in memory sub-system 140 in a synchronized manner. More specifically, in one embodiment, steps 220 and 230 are performed concurrently and the digital wave data and the text data generated

Confidential

is sent to memory sub-system 140 via switch 135 in alternate fashion such that a particular portion of the digital wave data is correlated with the corresponding portion of the text data as they are being stored as mixed data. In an exemplary embodiment, the present invention employs a buffering mechanism in

5   conjunction with switch 135 to handle timing delays that may arise during the voice recognition process (e.g., digital wave data is generated more quickly by HCE 120 than the corresponding text data is generated by VRE 130) to ensure that corresponding portions of digital wave data and text data is synchronized when it is stored in memory sub-system 140.

10

Referring next to Figure 2B, a diagram illustrating one embodiment of arrangement of corresponding portions of voice data and text data as stored in a portable digital recording and playback system 100 in accordance with the present invention is shown. In an exemplary embodiment as shown in Figure

15   2B, a voice input is converted into portions 261, 262 and 263 of digital wave data and corresponding portions 271, 272 and 273 of text data. These portions of digital wave data and text data are then stored in memory sub-system 140 as mixed data such that respective portions of digital wave data and text data are synchronized. More specifically, in one embodiment, the data portions are

20   stored in alternate fashion such that a particular portion of the digital wave data is correlated with the corresponding portion of the text data (e.g., text data portion 261 with digital wave data portion 271; text data portion 262 with digital wave data portion 272; text data portion 263 with digital wave data portion 273.)

As such, the present invention enables subsequent access and retrieval of the stored data to be performed efficiently and conveniently because the text data can be used to search for a desired portion of digital wave data, and vice versa, since the text data and digital wave data is synchronized. In one

5   embodiment, switch 135 is controlled based on phonetic group definitions of the text in the text data.

By performing real-time voice recognition on the voice input to generate text data, embodiments of the present invention eliminate the post-processing

10   that is typically required in prior art systems in order to derive text data from stored voice data. Moreover, since the text data is generated directly from the voice input in the present invention and not from highly compressed voice data as in the prior art, high quality voice-to-text conversion is achieved. In addition, since the present invention does not rely on the stored voice data to generate the

15   text data, the voice input can be subject to high compression and stored as digital wave data in accordance with the present invention to advantageously reduce the amount of memory required for storage without compromising the quality of the text data.

20   With reference next to Figure 3, a flow diagram 300 illustrating steps for performing playback using system 100 of Figure 1 in accordance with one embodiment of the present invention is shown. In step 310, system 100 of Figure 1 retrieves the mixed data which comprises digital wave data and text data from memory sub-system 140.

In step 320, system 100 of Figure 1 decodes the digital wave data into audio data using decoder 150. In step 330, system 100 converts the audio data into an analog signal using D/A converter 160. In optional step 340, in one

5   embodiment, system 100 amplifies the analog signal. In step 350, system 100 generates a voice output corresponding to the voice input from the analog signal.

It is appreciated that the present invention provides a high quality voice output. More specifically, the voice output is based on the recorded voice input

10   (as digital wave data) and is a high fidelity reproduction thereof, and not based on a simulated voice generated using text data.

With reference still to Figure 3, in step 360, system 100 of Figure 1 displays the text data using display sub-system 180. More specifically, in one

15   embodiment, the digital wave data and the text data retrieved is sent to decoder 150 and display sub-system 180 via switch 145 in alternate fashion such that output of the digital wave data by audio output device 170 and display of the text data by display sub-system 180 is synchronized. As such, the present invention affords great convenience to the reviewer of the recorded voice and text. In one

20   embodiment, switch 145 is controlled based on phonetic group definitions of the text in the text data.

It is appreciated that embodiments of the present invention can operate for extended periods of time under battery power (e.g., disposable batteries,

rechargeable batteries) because components of system 100 (Figure 1) in accordance with the present invention do not consume power at a high rate. Thus, the present invention provides a digital recording and playback system which is operable under battery power and is portable and wherein high quality

5   text data is generated from a voice input without requiring post-recording conversion.

Moreover, it is appreciated that system 100 of Figure 1 in accordance with embodiments of the present invention does not require specialized circuit

10   components or extensive retrofitting of existing computer systems and other appliances, because the circuit elements required for its implementation are commonly used in today's electronic appliances and are fully compatible with existing computer systems and other appliances. As such, a portable, battery-powered digital recording and playback system which does not require post-

15   processing to generate high quality text data, and which is conducive to use with existing computer systems and other appliances is provided by the present invention.

It is further appreciated that although exemplary values and operational

20   details (e.g., compression ratio of HCE 120, voice recognition techniques used in VRE 130) for various components are given with respect to embodiments of the present invention described above, such values and details are illustrative only and can vary within the scope and spirit of the present invention.

The preferred embodiment of the present invention, a digital recording and playback system with built-in voice recognition capability for concurrent text generation, is thus described. While the present invention has been described in particular embodiments, it should be appreciated that the present invention

5   should not be construed as limited by such embodiments, but rather construed according to the below claims.

CLAIMS

What is claimed is:

5 ~~1.   A digital recording and playback system comprising:~~

~~an audio capturing device configured to receive a voice input;~~

~~a high compression encoder (HCE) coupled to said audio capturing~~

~~device and configured to generate digital wave data corresponding to said voice~~

~~input;~~

10 ~~a voice recognition engine (VRE) coupled to said audio capturing device~~

~~and configured to generate text data corresponding to said voice input;~~

~~a memory sub-system selectively coupled to said HCE and said VRE and~~

~~configured to store said digital wave data and said text data; and~~

~~wherein said HCE and said VRE are operable to concurrently generate~~

15 ~~said digital wave data and said text data in response to said voice input such that~~

~~said digital wave data and said text data can be stored in a synchronized~~

~~manner.~~

~~2.   The system as recited in Claim 1 further comprising a first switch~~

20 ~~coupled between said HCE and said memory sub-system and also between said~~

~~VRE and said memory sub-system, said first switch configured to couple one of~~

~~said HCE and said VRE to said memory sub-system and to simultaneously~~

~~decouple the other one of said HCE and said VRE from said memory sub-~~

~~system.~~

25

3. ~~The system as recited in Claim 1 further comprising:~~

~~a decoder selectively coupled to said memory sub-system and configured to decode said digital wave data into decoded audio data;~~

~~a digital-to-analog (D/A) converter coupled to said decoder and configured~~
5 ~~to convert said decoded audio data into an analog signal; and~~

~~an audio output device coupled to said D/A converter and configured to render a voice output corresponding to said voice input from said analog signal.~~

4. ~~The system as recited in Claim 3 further comprising an amplifier~~
10 ~~coupled between said D/A converter and said audio output device and configured to amplify said analog signal.~~

5. ~~The system as recited in Claim 3 further comprising a display sub-system selectively coupled to said memory sub-system and configured to display~~
15 ~~said text data.~~

6. ~~The system as recited in Claim 5 further comprising a second switch coupled between said decoder and said memory sub-system and also between said display sub-system and said memory sub-system, said second~~
20 ~~switch configured to couple one of said decoder and said display sub-system to said memory sub-system and to simultaneously decouple the other one of said decoder and said display sub-system from said memory sub-system.~~

7. ~~The system as recited in Claim 5 wherein said display sub-system~~
25 ~~comprises a liquid crystal display (LCD).~~

8.   The system as recited in Claim 1 wherein said system is portable and battery-powered.

9.   The system as recited in Claim 1 wherein said memory sub-system

5   comprises semiconductor flash memory.

10.   The system as recited in Claim 1 wherein said VRE uses Hidden Markov Model (HMM) techniques to perform voice recognition.

10   11.   The system as recited in Claim 1 wherein said HCE is operable to achieve a compression ratio of two kilobits per second (2 kbit/s).

12.   The system as recited in Claim 2 wherein said first switch is controlled based on said text data.

15

13.   A method for audio recording and playback in a portable device, said method comprising the steps of:

a)   capturing a voice input;

b)   performing high compression encoding on said voice input to

20   generate digital wave data;

c)   performing voice recognition on said voice input to generate text data;

d)   storing said digital wave data and said text data in said portable device; and

25   wherein said steps b) and c) are performed concurrently to generate said digital wave data and said text data in response to said voice input such that said digital wave data and said text data can be stored in a synchronized manner.

14. The method as recited in Claim 13 wherein said step d) comprises the step d1) of alternately storing portions of said digital wave data and corresponding portions of said text data such that said digital wave data and said

5   text data is synchronized.

15. The method as recited in Claim 13 further comprising the steps of:

e)   retrieving said digital wave data from said portable device;

f)   decoding said digital wave data into decoded audio data;

10   g)   converting said decoded audio data into an analog signal; and

h)   generating a voice output corresponding to said voice input from said analog signal.

16. The method as recited in Claim 15 further comprising the step of

15   amplifying said analog signal.

17. The method as recited in Claim 15 further comprising the steps of:

i)   retrieving said text data from said portable device; and

j)   displaying said text data.

20

18. The method as recited in Claim 17 wherein said step e) comprises the step of retrieving portions of said digital wave data from said portable device and said step i) comprises the step of retrieving portions of said text data corresponding to said portions of said digital wave data from said portable

25   device, and wherein said steps e) and i) are performed alternately such that said retrieving of said digital wave data and said text data is synchronized.

Confidential

19.   The method as recited in Claim 17 wherein said step j) comprises the step of displaying said text data on a liquid crystal display (LCD).

5

20.   The method as recited in Claim 13 wherein said portable device is battery-powered.

21.   The method as recited in Claim 13 wherein said step d) comprises the step of storing said digital wave data and said text data in semiconductor flash memory within said portable device.

10

22.   The method as recited in Claim 13 wherein said step c) comprises the step of performing voice recognition on said voice input to generate text data using Hidden Markov Model (HMM) techniques.

15   23.   The method as recited in Claim 13 wherein said high compression encoding achieves a compression ratio of two kilobits per second (2 kbit/s).

24.   The method as recited in Claim 14 wherein said step d1) is controlled based on said text data.

20

25.   A digital recording and playback system comprising:
an audio capturing means for receiving a voice input;
a high compression encoding means coupled to said audio capturing means for generating digital wave data corresponding to said voice input;
25   a voice recognition means coupled to said audio capturing means for generating text data corresponding to said voice input; and

a storage means selectively coupled to said high compression encoding means and said voice recognition means for storing said digital wave data and said text data, wherein said high compression encoding means and said voice recognition means are operable to concurrently generate said digital wave data

5   and said text data in response to said voice input such that said digital wave data and said text data can be stored in a synchronized manner.

26.   The system as recited in Claim 25 further comprising a first switching means coupled between said high compression encoding means and

10   said storage means and also between said voice recognition means and said storage means, said first switching means for coupling one of said high compression encoding means and said voice recognition means to said storage means while simultaneously decoupling the other one of said high compression encoding means and said voice recognition means from said storage means.

15

27.   The system as recited in Claim 25 further comprising:
a decoding means selectively coupled to said storage means for decoding said digital wave data into decoded audio data;
a digital-to-analog (D/A) converting means coupled to said decoding

20   means for converting said decoded audio data into an analog signal; and
an audio output means coupled to said D/A converting means for generating a voice output corresponding to said voice input from said analog signal.

25   28.   The system as recited in Claim 27 further comprising an amplifying means coupled between said D/A converting means and said audio output means for amplifying said analog signal.

29. The system as recited in Claim 27 further comprising a display means selectively coupled to said storage means for displaying said text data.

5   30. The system as recited in Claim 29 further comprising a second switching means coupled between said decoding means and said storage means and also between said display means and said storage means, said second switch for coupling one of said decoding means and said display means to said storage means while simultaneously decoupling the other one of said decoding means and said display means from said storage means.

63.   A recording and playback system comprising:

an audio capturing device configured to receive an analog input;

an encoder coupled to said audio capturing device and configured to

15   generate a digital signal based on said analog input; and

a recognition engine coupled to said audio capturing device and configured to generate text data based on said analog input, wherein said encoder and said recognition engine simultaneously generate said digital signal and said text data such that said digital signal and said text data can be provided

20   in a synchronized manner.

64.   The system as recited in Claim 63 further comprising a first switch coupled between said encoder and a memory sub-system and also between said recognition engine and said memory sub-system, said first switch configured to

25   couple one of said encoder and said recognition engine to said memory sub-system and to simultaneously decouple the other one of said encoder and said recognition engine from said memory sub-system.

65.   The system as recited in Claim 64 further comprising:

a decoder selectively coupled to said memory sub-system and configured to decode said digital signal into decoded audio data;

5   a digital-to-analog (D/A) converter coupled to said decoder and configured to convert said decoded audio data into an analog signal; and

an audio output device coupled to said D/A converter and configured to render a voice output corresponding to said analog input from said analog signal.

10   66.   The system as recited in Claim 65 further comprising an amplifier coupled between said D/A converter and said audio output device and configured to amplify said analog signal.

67.   The system as recited in Claim 65 further comprising a display sub-

15   system selectively coupled to said memory sub-system and configured to display said text data.

68.   The system as recited in Claim 67 further comprising a second switch coupled between said decoder and said memory sub-system and also

20   between said display sub-system and said memory sub-system, said second switch configured to couple one of said decoder and said display sub-system to said memory sub-system and to simultaneously decouple the other one of said decoder and said display sub-system from said memory sub-system.

25   69.   The system as recited in Claim 67 wherein said display sub-system comprises a liquid crystal display (LCD).

70.   The system as recited in Claim 63 wherein said system is portable and battery-powered.

71.   The system as recited in Claim 64 wherein said memory sub-

5   system comprises semiconductor flash memory.

72.   The system as recited in Claim 63 wherein said recognition engine uses Hidden Markov Model (HMM) techniques to perform recognition.

10   73.   The system as recited in Claim 63 wherein said encoder is operable to achieve a rate of two kilobits per second (2 kbit/s).

74.   The system as recited in Claim 64 wherein said first switch is controlled based on said text data.

15

75.   A method for audio recording and playback in a portable device, said method comprising the steps of:

        a)   capturing a first analog signal;

        b)   encoding said first analog signal to generate a digital signal;

20   c)   performing recognition on said analog signal to generate text data;

        wherein said b) and c) are performed simultaneously to generate said digital signal and said text data in response to said first analog signal such that said digital signal and said text data can be stored in a synchronized manner in a memory device.

25

76.   The method as recited in Claim 75 further comprising:

Confidential

alternately storing portions of said digital signal and corresponding portions of said text data such that said digital signal and said text data is synchronized.

5   77.   The method as recited in Claim 75 further comprising the steps of:

d)   decoding said digital signal into decoded audio data;

e)   converting said decoded audio data into a second analog signal; and

f)   generating a voice output corresponding to said first analog signal

10   from said second analog signal.

78.   The method as recited in Claim 77 further comprising the step of amplifying said second analog signal.

15   79.   The method as recited in Claim 77 further comprising the steps of:

g)   displaying said text data on a display device coupled to said portable device.

80.   The method as recited in Claim 79 wherein said display device is a

20   liquid crystal display (LCD).

81.   The method as recited in Claim 75 wherein said portable device is battery-powered.

25   82.   (New) The method as recited in Claim 75 wherein said memory device is a flash memory coupled to said portable device.

83.   (New) The method as recited in Claim 75 wherein a Hidden Markov Model (HMM) technique is used for said recognition.

84.   (New) The method as recited in Claim 75 wherein said encoding is
5   performed at a rate of substantially two kilobits per second (2 kbit/s).

DIGITAL RECORDING AND PLAYBACK SYSTEM

WITH VOICE RECOGNITION CAPABILITY

FOR CONCURRENT TEXT GENERATION

5 ABSTRACT OF THE INVENTION

A recording and playback system is provided. The system includes an audio capturing device configured to receive an analog input and an encoder coupled to the audio capturing device configured to generate a digital signal based on the analog input. The system further includes a recognition engine

10 coupled to the audio capturing device and configured to generate text data based on the analog input, wherein the encoder and the recognition engine simultaneously generate the digital signal and the text data such that the digital signal and the text data can be provided in a synchronized manner.

15 A digital recording and playback system with built-in voice recognition capability for concurrent text generation. In one embodiment, the system comprises an audio capturing device configured to receive a voice input. The system also comprises a high compression encoder (HCE) coupled to the audio capturing device and configured to generate digital wave data corresponding to

20 the voice input, as well as a voice recognition engine (VRE) coupled to the audio capturing device and configured to generate text data corresponding to the voice input. In this embodiment, the HCE and VRE are selectively coupled to a memory sub-system which is configured to store the digital wave data and the text data. In this embodiment, the VRE performs voice-to-text conversion using

25 the high quality audio input signal rather than highly compressed voice data so that high quality conversion is achieved. In this embodiment, the HCE and the VRE are operable to concurrently generate the digital wave data and the text

data in response to the voice input such that the digital wave data and the text data can be stored in the memory sub-system in a synchronized manner. As such, this embodiment of the present invention provides recording capability wherein text data is generated from a voice input without requiring post-recording conversion. In a specific embodiment, the present invention includes the above and wherein the system is battery-powered and is portable.